International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 10, Issue 6, June 2021)

An Analysis of Udergraduate Students Parameters for Predicting passing Ratio using Artificial Intelligence

Akhilesh Kumar[1], Nitish Kumar[2], Deepak Kumar Choudhary[3], Shweta Kumari[4]

[1,3,4]*Department of Computer Science and Engineering, Darbhanga College of Engineering, Darbhanga*
[2]*Department of Electrical and Electronics Engineering, Darbhanga College of Engineering, Darbhanga*

***Abstract*— Identification of different factors which affects a student's learning behavior and performance during academic career. Analyzing student mental issues and low academic performances is a complex task in the current education sector. In current system it is difficult to track the student's behavior and characteristics. There is no automation or tool which predicts or shows how to improvise the student's academic performances. Machine learning (ML) algorithms is use to better understand learners' learning is popular in the educational community. Studies generally adopt supervised learning to build a learning model with which to predict students' performance or to identify at-risk students. This paper presents the review of machine learning techniques for examining successful attributes of undergraduate students.**

***Keywords*— AI, Machine Learning, Examining, Attributes, Undergraduate Students, Performance.**

I. INTRODUCTION

The meaningful information regarding student learning and improve the interpretation of the finding, the present research adopts the self-regulation theory as a framework and employs both supervised and unsupervised learning techniques to explore which attributes are critical for a student during his/her learning. A major problem for student affairs management is the contradiction between the limited energy of student counselors and the diversity of student behaviors, which results in many potential problem students losing the opportunity for early intervention. Since the beginning of the 21st century, the rapid development of information technology in education and the construction of digital campuses has made it possible for student counselors to conduct quantitative analysis of student school behaviors, especially to provide early warning to students who may have problems, so that the contradiction could be alleviated by applying the analysis and early warning methods. As contemporary college students who grew up in the Internet era, their daily life, learning and thinking are deeply influenced by the Internet. This provides us with the possibility to understand their campus network behavioral characteristics through big data.

How to mine useful information for student counselor from massive data in the explosive growth of data categories and data scales, is a challenge for current student counselor, also an important opportunity to conduct work by new means. In today's life, emotional wellbeing is an important aspect to be analyzed particularly for school going children. If not properly analyzed, the emotional difficulties may hinder academic, personal and social growth, resulting in a lifetime of difficulties for such individuals. Only when a child is emotionally secure and content, he/she gives their best to every challenge they face. Schools should create a progressive learning environment where academic achievement is related not only to successful learning strategies, but also to good mental well-being [4].

E-learning paradigm is the most successful part in the arena of Educational Data Mining (EDM). People prefer to get a lot of learning materials in the e-learning platforms rather than physical classrooms. In all educational institutions, it becomes a part of students' academic activities to enroll in any of their prescribed online courses. Moodle is a popular management system which paves way for students and teachers to take learning in a comfort zone. A number of parameters such as assignment submission, assessment criteria, number of clicks made by the learners on a single day, enrollment for the exams etc. can be achieved easily in the Moodle platform. Student dataset of an educational institution is taken for the proposed work. Courses selected by the learners are categorized as low, medium and high level courses with respect to its complexity based on few required parameters [6]. The prediction of SGPA and CGPA is beneficial to university students. Students will easily get an estimate of their final outcome from this project. As a result, the students will be able to brace themselves for a successful outcome. Students pass the day by participating in a variety of events. Students use social media sites such as Facebook, Instagram, and Twitter. They engage in various hobbies such as playing mobile games, listening to music, among others. As a result, they were able to move several times with these tasks.

As a result, if a student spends so much time doing any of those things, she will not be able to achieve a successful grade because of the experiment; students can develop a research routine or guideline that they can apply to their other tasks. Additionally, students' behaviors will forecast their outcomes [7]. Student semester grade point average (GPA) is measure of student success to take into account the temporal effects in student success. The findings highlight the student performance based on their demographic status and use of university resources such as financial aid. College campuses should not only increase current resources but also raise awareness of current resources and make them more accessible (e.g., easier to apply or automatic applications). This is especially important for some demographics such as Hispanic first-generation students. Background: Higher education institutions are facing retention and graduation problems. One way to improve this is by understanding why students are not academically successful [10].

II. Literature Survey

C. -Y. Ko et al.,[1] This study applies supervised and unsupervised machine learning (ML) techniques to discover which significant attributes that a successful learner often demonstrated in a computer course. Background: Students often experienced difficulties in learning an introduction to computers course. This research attempts to investigate how successful students regulate their learnings in this course. The answer to these questions will provide teachers with useful information to better comprehend how students learn and which strategies are effective in learning. Research Questions: 1) Which algorithm in supervised learning is the best one for predicting students' final performance? and 2) What attributes are key to succeed in this course? Methodology: Seven supervised ML algorithms and ensembles are conducted to compare the performance of classifiers regarding the levels of accuracy, precision, and sensitivity. The association rule and clustering are also employed to discover the key attributes for successful students. Because the present study used a convenience sample for data analysis, the number of students in each cluster was a potential limitation. Findings: The results show that Naïve Bayes is the most appropriate one for predicting students' final performance. The measure in accuracy and sensitivity of this classifier achieves 83.26% and 92.88%, respectively.

In addition, the association rule indicates that "to make sure keep up with the weekly progress in the class" and self-efficacy beliefs play important roles on final performances for learners. The clustering findings reveal similar results.

P. R. Tlalpan et al.,[2] The objective of this intervention project was to evaluate the impact of learning activities strengthened with Augmented Reality (AR) in Mathematics and the generation of positive and negative emotions in students who attended the third semester of the ITESM high school, Mexico City campus, where there had been a high percentage of failed students taking the subject "Algebraic and Transcendent Functions." Evaluating the academic performance and the students' emotions before and after carrying out the activities, we obtained a p-value of 0.0014 in the non-parametric Wilcoxon test, indicating a statistically significant change in acquired learning. AR activities were implemented using different pedagogical strategies. We attained statistically significant results, increasing students' grades by 68%, achieving a grade point average of 93.47 out of 100 in the topics covered. After the intervention, negative emotions decreased by 84.2%, and positive emotions increased by 85%. We concluded that AR is an excellent technological tool to support remote mathematics classes, changing how students interact. The results also indicated that changes in pedagogical strategies are essential to the successful implementation of the remote teaching modality; also, using various stimuli changed the emotions experienced by the students.

J. M. Campos et al.,[3] This article describes a data science challenge-based learning experience introduced to non-IT second-year engineering students. The methodology proposed in this article was successful in the short introductory course. The students presented well-considered, practical solutions to two challenges of a different nature derived from excellent, quality data processing. The students used free-access databases from Airbnb and Johns Hopkins University to tackle both challenges. Although the students' data analysis methods corresponded more to data analytics, the two student teams incorporated Machine Learning techniques and exceeded our initial expectations.

M. Varsha, et al.,[4] aim to explore the meaning, situation, and parties associated with happiness in children and to deliver the best tool to detect the state of the children's mental health regarding whether they are happy or sad and based on machine learning techniques.

H. Yang et al.,[5] This full research paper presents a systematic analysis of 10 years' student performance data of Computer Science (CS) majors at San Francisco State University, a public 4-year degree-granting university, aiming to address the ongoing challenges of early dropouts and low graduation rate. The main objective is two-fold: (1) gain a comprehensive understanding of how the existing curriculum has been supporting (or hindering) students' progress towards graduation; and (2) suggest data-informed curricular changes. To this end, we utilize both explorative statistical analysis and data mining/machine learning approaches to first learn how individual courses and the prescribed course sequences influence a student's dropout/graduation status, and then build machine learning models to interpret/validate the observed interdependency among key courses in the current curriculum. Such patterns/models are consequently utilized to suggest impactful curricular changes towards reducing early dropouts and improving the overall student success as measured by graduation with a CS degree. One main finding of this research is that a successful CS student needs to excel in both critical thinking and core CS skills.

S. Subha et al.,[6] This work aims at comparing the accuracy obtained by different Machine Learning (ML) algorithms for the given data set. Among the input, 60% data is taken as training data set and 40% is considered as the testing data. Confusion matrix has been generated using machine learning algorithms namely Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Logistic regression(LR), and Random Forest. Important metrics such as recall, precision, F-measure and support have been computed and the classification report has been generated for each algorithm. Experimental results produce 97.5%accuracy with Random forest algorithm and it is revealed to be high when compared to the other specified ML algorithms. Thereby, these results are used for analyzing the performance level of the students.

M. Saifuzzaman et al.,[7] The Authors will now see machine learning in Python being used all over the place. After that, The Authors created a smart SGPA and CGPA prediction project, as well as the results on students. The findings are predicted using the Nave Byes algorithm. The Nave Byes algorithm is a simple but effective prediction algorithm. It is a machine learning algorithm as well. As a result, students will be given an estimate of their final exam scores. They can prepare them to make a good result by following the routine of the SGPA & CGPA prediction project.

R. Alamri et al.,[8] Successful prediction of student performance has significant impact to many stakeholders, including students, teachers and educational institutes.

In this domain, it is equally important to have accurate and explainable predictions, where accuracy refers to the correctness of the predicted value, and explainability refers to the understandability of the prediction made. In this systematic review, we investigate explainable models of student performance prediction from 2015 to 2020. We analyze and synthesize primary studies, and group them based on nine dimensions. Our analysis revealed the need for more studies on explainable student performance prediction models, where both accuracy and explainability are properly quantified and evaluated.

L. M. Cruz Castro et al.,[9] This study supports the characterization of four different student profiles demonstrating differences in their performance at the beginning of the semester. From these four profiles, two of them show a subsequent differential progression besides their similarity at the beginning of the semester. In this particular case, troubleshooting and debugging appear as a relevant competency when distinguishing these two learners' groups. These findings suggest that previous knowledge or exposure to different practices can result in different progressions of more complex computational practices, emphasizing the relevance of troubleshooting and debugging as a practice required for a successful and timely progression on the acquisition of other computational thinking practices.

F. Marbouti et al.,[10] this study presents, demographic information and past academic records were analyzed to understand patterns of student success. Methodology: A cluster analysis was conducted to understand groups of students based on academic performance and demographic information. Examples of these factors are enrollment status, financial status, first-generation status, housing status, and transfer status. For the purpose of getting more accurate results, the students were separated into two different groups according to their admission status: 1) freshman and 2) transfer. Findings: The results indicate Hispanic, first-generation, low-income students are not likely to apply for financial aid although they are eligible. They have lower GPA and take fewer units per semester than other students. This can cause delayed graduation and accumulating more debt.

R. Ajoodha et al.,[11] attempt to provide a data-driven solution to the data-congested environment of attributes related to student success and contribute towards preventing the increased dropout rates at South African higher education institutions. One of the most significant discussions in higher education is student attrition in their first year of study. Student career guidance is an area that requires investigation in light of high attrition rates at university.

Recent developments in data analytics, and the analysis of large data sets have enabled the production of powerful predictive models. This paper highlights how a predictive model can assist students, with an interest in Science to develop a skill profile required to be successful in their undergraduate Science programme. This is achieved by identifying the difference between the necessary skills required to be successful in a science programme (derived using data driven approaches) from the current learner's skill profile (derived from the learners' performance in assessments).

Đ. Pašić et al.,[12] Transition from high school to university is not successful for all students. Before enrolling in a program, the Admission Office tries to help those students to decide whether that program is best suited for them. They do so by using collected empirical data. The challenge is to classify which students would successfully finish the program and which would not. The students who are most likely not to finish the program successfully should be warned that their decision is not necessarily the best decision for them, and they should consider some other possibilities. This paper proposes one possible criterion for the given challenge: we will develop a machine learning model with the collected data from high schools which the students have attended prior to Algebra and calculate the probability of not finishing the program successfully. Students who would be classified by the model as unsuccessful will receive recommendation for another study program.

III. Disadvantages And Gap Analysis

Disadvantages-

- System predicts student performance based on the student behaviors. Normal or unauthorized Data-set not compatible for class results prediction.
- Used more irrelevant parameters for student performance prediction such as father income, mother income, qualification etc.

Gap Analysis-

- To build a real world application this is useful for colleges and students to improvise performance.
- To develop a system which predicts the relationship between students behavior and performance using the educational data-set.
- Proposed project is a student's behavior analysis and prediction and management system which is meant for educational institute.
- System aim is to reduce the number of students with poor performances.
- System makes use of "Association Learning" to discover the educational patterns using data science algorithms.

IV. Conclusion

Identification of different factors which affects a student's learning behavior and performance during academic career. Analyzing student mental issues and low academic performances is a complex task in the current education sector. System uses data. science technique called as "Association Learning" to find the patterns. Further work to apply different algorithms can be used to find the students behaviour patterns and algorithms can be compared to identify the algorithm with better results.

REFERENCES

[1] C. -Y. Ko and F. -Y. Leu, "Examining Successful Attributes for Undergraduate Students by Applying Machine Learning Techniques," in IEEE Transactions on Education, vol. 64, no. 1, pp. 50-57, Feb. 2021, doi: 10.1109/TE.2020.3004596.

[2] P. R. Tlalpan, "The Impact of Augmented Reality on Student Learning and Emotions in Mathematics," 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop, 2021, pp. 1-6, doi: 10.1109/IEEECONF53024.2021.9733781.

[3] J. M. Campos, E. A. Lozano, J. Urzúa and J. G. Calderón, "Challenge Based Learning: A Fast Track To Introduce Engineering Students To Data Science," 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop, 2021, pp. 1-6, doi: 10.1109/IEEECONF53024.2021.9733765.

[4] M. Varsha, M. Ramya, C. C. Sobin, N. Subheesh and J. Ali, "Assessing Emotional Well-being of Students using Machine Learning Techniques," 2021 19th OITS International Conference on Information Technology (OCIT), 2021, pp. 336-340, doi: 10.1109/OCIT53463.2021.00073.

[5] H. Yang, T. W. Olson and A. Puder, "Analyzing Computer Science Students' Performance Data to Identify ImpactfulCurricular Changes," 2021 IEEE Frontiers in Education Conference (FIE), 2021, pp. 1-9, doi: 10.1109/FIE49875.2021.9637474.

[6] S. Subha and S. B. Priya, "Comparative Analysis of Supervised Machine Learning Algorithms for Evaluating the Performance Level of Students," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2021, pp. 1-10, doi: 10.1109/I-SMAC52330.2021.9640798.

[7] M. Saifuzzaman, M. Parvin, I. Jahan, N. N. Moon, F. Narin Nur and S. F. Shetu, "Machine Learning Approach to Predict SGPA and CGPA," 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), 2021, pp. 211-216, doi: 10.1109/ICAICST53116.2021.9497847.

[8] R. Alamri and B. Alharbi, "Explainable Student Performance Prediction Models: A Systematic Review," in IEEE Access, vol. 9, pp. 33132-33143, 2021, doi: 10.1109/ACCESS.2021.3061368.

[9] L. M. Cruz Castro, A. J. Magana, K. A. Douglas and M. Boutin, "Analyzing Students' Computational Thinking Practices in a First-Year Engineering Course," in IEEE Access, vol. 9, pp. 33041-33050, 2021, doi: 10.1109/ACCESS.2021.3061277.

[10] F. Marbouti, J. Ulas and C. -H. Wang, "Academic and Demographic Cluster Analysis of Engineering Student Success," in IEEE Transactions on Education, vol. 64, no. 3, pp. 261-266, Aug. 2021, doi: 10.1109/TE.2020.3036824.

[11] R. Ajoodha, S. Dukhan and A. Jadhav, "Data-Driven Student Support for Academic Success by Developing Student Skill Profiles," 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), 2020, pp. 1-8, doi: 10.1109/IMITEC50163.2020.9334109.

[12] Đ. Pašić and D. Kučak, "Machine learning model for detecting high school students as candidates for drop-out from a study program," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, pp. 1140-1144, doi: 10.23919/MIPRO48935.2020.9245405.